Máquinas Tendenciosas: Até onde vai o impacto social nas decisões tomadas por softwares inteligentes?

Trabalho da Cadeira de Introdução à Ciência dos Dados

Autores: Lerisson Freitas(lff3);

Matheus Raz (mrol)

Recife, 28 de Outubro de 2018

Introdução

Esse trabalho tem como objetivo resumir o conteúdo abordado em 4 artigos jornalísticos para uma melhor compreensão destas. Estes rodeiam sobre a temática de "Inteligências Artificiais Racistas". Os artigos lidos podem ser encontrados abaixo:

- **Rise of the racist robots** – how AI is learning all our worst impulses. Stephen Buranyi. The Guardian, 2017.https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses
- **Machine Bias:** There's software used across the country to predict future criminals. And it's biased against blacks. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. ProPublica, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- **Google says sorry for racist auto-tag in photo app**. Jana Kasperkevic. The Guardian, 2015.https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app
- **Facebook (Still) Letting Housing Advertisers Exclude Users by Race Julia Angwin**, Ariana Tobin and Madeleine Varner. ProPublica, 2017. https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

Resumo

O racismo infelizmente é uma questão antiga que perdura até os dias atuais, e inerente dos humanos, seja por valores adquiridos durante a formação de seu caráter ou por “preconceitos” de tal temática. Hoje com os avanços tecnológicos, há várias discussões éticas da comunidade a respeito de terceirizar as decisões ou tomar estas com o auxílio da Inteligência Artificial, com o intuito de alcançar decisões corretas e imparciais e ter um julgamento melhor das coisas.

O problema abordado nos dois primeiros artigos são voltados para um viés de julgamento quanto aos índices de criminalidade nos USA, onde a polícia local fez uso de programas terceirizados de empresas privadas, que a partir de dados históricos sobre crimes recorrentes e locais onde eles ocorrem, tentavam predizer a probabilidade de uma pessoa cometer algum crime, porém, alguns problemas começaram a ser identificados quanto às diversas variáveis que os programas levam em conta, e outras que apenas a concepção humana seria capaz de julgar. No julgamento comum como sempre foi feito pelo ser humano, sempre houve uma tendência preconceituosa quanto a classe e a cor de um indivíduo, e o motivo do uso de tais recursos seria exatamente obter uma forma mais justa de se julgar qualquer potencial criminoso. Contudo, os dados históricos já possuíam características que propunham um grau de preconceito estabelecido por julgamentos humanos, e os programas começaram a replicar tais falhas executadas no passado apenas automatizando algo que era aplicado, como o próprio termo na área já indica “Garbage In, Garbage Out”, indicando que os algoritmos apenas reproduzem apenas os conceitos que ele aprende, se passam dados que possuem um viés tendencioso, o algoritmo também será tendencioso.

Já dizia um cientista da computação que é impossível saber o quão amplamente adotada IA é agora, mas se sabe que não podemos voltar atrás. E mesmo com todo investimento ainda é uma área de vanguarda que requer todo um estudo de como considerar as melhores variáveis para deixar seu julgamento idôneo. Kristian Lum já dizia “Se você não tomar cuidado, corre o risco de automatizar exatamente os mesmos preconceitos que esses programas devem eliminar".

Como mostra uma pesquisa vista no artigo 2 em 2009, Brennan e dois colegas publicaram um estudo de validação que descobriu que o risco de pontuação de ser reincidente da Northpointe tinha uma taxa de precisão de 68% em uma amostra de 2.328 pessoas. O estudo também descobriu que a pontuação era um pouco menos preditiva para homens negros do que para homens brancos - 67% contra 69%. Ele não examinou as disparidades raciais, além disso, incluiu se alguns grupos eram mais propensos a ser erroneamente rotulados como de maior risco.

Indo de desencontro com o que foi apresentado acima, o google não se mostrou ser resistente a mudanças e reconheceu seu erro cedo, afirmando que não faz parte da política da

empresas posicionamento de grupos específicos, e sim defende a pluralidade de público. Este pronunciamento se deu no caso do google fotos, especificamente na auto sugestão de tags nas imagens. Tal situação se deu quando uma foto de negros foi classificada como gorilas. Devido ao desconhecimento do algoritmo de tal classificação, e este já se mostrava com dificuldade em classificar pessoas por suas semelhanças com outros animais. Já dizia o própria empresa: “se você excluir uma tag incorreta, nosso algoritmo aprende com esse erro e terá um desempenho melhor no futuro. O processo de marcação é completamente automatizado - nenhum ser humano jamais verá suas fotos para marcá-las” , ou seja, o algoritmo é retroalimentado e aprende com seus próprios erros.

Por fim, a política do facebook quanto a discriminação aparenta ser rígida, contudo foi visto que existe uma brecha que permite aos anunciantes, excluírem os negros, hispânicas e outras “afinidades étnicas” da exibição de anúncios, ou também por disposição geográfica desses raças. A empresa diz que conseguiu contornar essa problemática, mas e uma questão que deve ser melhorada.

Concluímos então que apesar do grande poder que o uso desses programas inteligentes trazem nas tomadas de decisão, há diversos contextos em que seu uso não leva em conta variáveis diretamente dependentes de vieses humanos, como no caso do julgamento de potenciais criminosos, ou que suas taxas de erro possam gerar resultados que firam os direitos humanos e impactar preconceitos entranhados na sociedade causando revolta e má satisfação dos usuários.